



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/944,332	08/30/2001	Liqin Shen	JP920000191US1 (590.079)	1865
35195	7590	06/21/2005	EXAMINER	
FERENCE & ASSOCIATES 409 BROAD STREET PITTSBURGH, PA 15143			HAN, QI	
			ART UNIT	PAPER NUMBER
			2654	
DATE MAILED: 06/21/2005				

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)
	09/944,332	SHEN ET AL.
	Examiner	Art Unit
	Qi Han	2654

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
 - If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
 - If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
 - Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) Responsive to communication(s) filed on ____.
- 2a) This action is FINAL. 2b) This action is non-final.
- 3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) Claim(s) 1-19 is/are pending in the application.
- 4a) Of the above claim(s) ____ is/are withdrawn from consideration.
- 5) Claim(s) ____ is/are allowed.
- 6) Claim(s) 1-19 is/are rejected.
- 7) Claim(s) ____ is/are objected to.
- 8) Claim(s) ____ are subject to restriction and/or election requirement.

Application Papers

- 9) The specification is objected to by the Examiner.
- 10) The drawing(s) filed on ____ is/are: a) accepted or b) objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) All b) Some * c) None of:
 1. Certified copies of the priority documents have been received.
 2. Certified copies of the priority documents have been received in Application No. ____.
 3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- | | |
|---|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413) |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | Paper No(s)/Mail Date: ____ |
| 3) <input checked="" type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
Paper No(s)/Mail Date <u>05/31/2002</u> . | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152) |
| | 6) <input type="checkbox"/> Other: ____ |

DETAILED ACTION

Information Disclosure Statement

1. The references listed in the Information Disclosure Statement submitted on 05/31/2002 and 01/18/2005 have been considered by the examiner (see attached PTO-1449).

Specification

2. The disclosure is objected to because of the following:
 - a. On page 6, line 7, the content “length (S)-N is N(N+1)/2” is unclear. Appropriate correction or explanation is required.
 - b. On page 10, lines 4-7, the disclosure recites “It can be seen from Table 1 that SBP A, B and C, the number of GAST notes, … reduces dramatically.” However, Table 1 shows that when the basic vocabulary increases from 3.6k to 4.3k under the condition SBP A+B+C, the average length and the number of GAST notes all increases (not reduces), which is in conflict with above statement in the specification. Appropriate correction or explanation is required.
 - c. On page 5, line 15, the term “ANWE” lacks an antecedent definition or description, since it is not commonly used term in the art. Appropriate correction is required.

Claim Rejections - 35 USC § 112

The following is a quotation of the second paragraph of 35 U.S.C. 112:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

3. Claims 1 and 10 rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

Regarding claims 1 and 10, the limitation “a **cleaned** corpus” lacks a clear scope in the claim, since the specification does not specifically describe or clearly define what level or type of “cleanliness” is for a corpus, and the limitation are not commonly accepted terms in the art, which leads to the claimed limitation to be indefinite.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

4. Claims 1-3, 6-12 and 15-19 are rejected under 35 U.S.C. 103(a) as being unpatentable over Wang et al. (US 6,904,402 B1) hereinafter referenced as Wang, in view of Razin et al. (US 6,098,034) hereinafter referenced as Razin.

As per **claim 1**, as best understood in view of the rejection under 112, 2nd (see above), Wang discloses system and iterative method for lexicon, segmentation and language model joint optimization (title), comprising:

“segmenting a cleaned corpus to form a segmented corpus”, (Fig. 5 and column 9, lines 36-44, ‘segmentation’, ‘the received corpus is built’, ‘pre-processed to remove some obvious illogical words (so as to provide cleaned corpus)’);

“splitting the segmented corpus to form sub strings, and counting the occurrences of each sub strings appearing in the corpus” (column 1, lines 45-60, ‘a textual corpus is dissected (interpreted as split) into a plurality of items (sub strings)’ and ‘counts the number of occurrences of a particular item (word, character, etc.)’); and

Even though Wang further suggests that ‘the items of the corpus’ having low occurrence frequency ‘may be pruned’, Wang does not expressly discloses “filtering out false candidates to output new words”. However, this feature is well known in the art as evidenced by Razin who, in the same field of endeavor, discloses method for standardizing phrasing in a document (title), comprising ‘filtering the preliminary list of extracted phrases (candidates) to create (output) a final list of extracted phrases (corresponding new words)’ (Fig. 2 and column 29, lines 30, lines 55-56). Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Wang by specifically providing filtering a set of extracted phrases and creating (output) final phrases list, as taught by Razin, for the purpose of obtaining extracted words constituting significant user phrases (Razin: column 2, lines 46-47).

As per **claim 2** (depending on claim 1), Wang in view of Razin further discloses “using punctuations, Arabic digits and alphabetic strings, or new words patterns to split the cleaned corpus”, (Razin, column 21, lines 10, ‘punctuation’; column 4, lines 26, ‘the usage of stop list’);

As per **claim 3** (depending on claim 1), Wang in view of Razin further discloses “using common vocabulary to segment the cleaned corpus”, (Razin: column 5, lines 36-45, ‘the dictionary of standard phrases (common vocabulary)’).

As per **claim 6** (depending on claim 1), Wang in view of Razin further discloses: “filtering out functional words” (Razin: column 4, lines 35-38, ‘stop list’, ‘semantically insignificant words (e.g., “and then about the”) (interpreted as functional words)’, which suggests that these words can be filtered out);

“filtering out those sub strings which almost always appear along with a longer sub strings” (Razin: column 9, lines 52, ‘eliminates from the phrase list otherwise-significant phrases that are nested within other significant phrases... removes from the final phrase list minimal content words dangling at the beginning or end of preliminary user-specific phrases’, which reads on the claim); and

”filtering out those sub strings for which the occurrence is less than a predetermined threshold”, (Razin: column 2, lines 10-13, ‘each node of tree is associated with a record of the number of occurrence of the word sequence at that node, where the number of occurrence exceeds the required threshold’, which reads on the claimed limitation).

As per **claim 7** (depending on claim 1), Wang in view of Razin further discloses “using pre-recognized functional words as segment boundary patterns”, (Razin: column 4, lines 35-38, ‘stop list’, ‘semantically insignificant words (e.g., “and then about the”) (interpreted as functional words)’).

As per **claim 8** (depending on claim 3), the rejection is based on the same reason described for claim 7 because the claim recites the same or similar limitation(s) as claim 7.

As per **claim 9** (depending on claim 3), the rejection is based on the same reason described for claim 6 because the claim recites the same or similar limitation(s) as claim 6.

As per **claims 10-12 and 15-18**, they recite an automatic new word extraction system. The rejection is based on the same reason described for claims 1-3 and 6-9, respectively, because the claims recite the same or similar limitation(s) as claims 1-3 and 6-9, respectively.

As per **claim 19**, it recites a program storage device readable by machine. The rejection is based on the same reason described for claim 1, because the claim recites the same or similar limitations as claim 1.

5. Claims 4-5 and 13-14 are rejected under 35 U.S.C. 103(a) as being unpatentable over Wang in view of Razin as applied to claims 1 and 10, and further in view of Hui (IDS: "Color Set Size Problem with Applications to String Matching," Proc. of 2nd Symposium on Combinatorial Pattern Matching, 1992, pp. 230-243).

As per **claim 4** (depending on claim 1), even Wang in view of Razin further discloses using suffix tree (i.e. atomic suffix tree—AST) (Wang: column 1, line 42; Razin: column, 2, line 3), Wang in view of Razin does not expressly disclose "using a GAST". However, the feature is well known in the art as evidenced by Hui who teaches 'the concept of suffix tree can be extended' and 'this extension is called the Generalized suffix tree (GST)(corresponding to GAST)' (Hui, page 237, first paragraph). Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Wang in view of Razin by specifically providing using extended suffix tree (GST or GAST), for the purpose of storing more than one input strings (Hui: page 237, first paragraph).

As per **claim 5** (depending on claim 4), Wang in view of Razin and Hui further discloses the tree “implemented by limiting length of sub strings”, (Razin: column 14, lines 34-35, ‘length less than or equal to Smax’).

As per **claim 13** (depending on claim 10), the rejection is based on the same reason described for claim 4 because the claim recites the same or similar limitation(s) as claim 4.

As per **claim 14** (depending on claim 10), the rejection is based on the same reason described for claim 5 because the claim recites the same or similar limitation(s) as claim 5.

Conclusion

6. Please address mail to be delivered by the United States Postal Service (USPS) as follows:

Mail Stop _____
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

or faxed to: (703) 872-9306, (for formal communications intended for entry)
Or: (703) 872-9306, (for informal or draft communications, and please label "PROPOSED" or "DRAFT")

If no Mail Stop is indicated below, the line beginning Mail Stop should be omitted from the address.

Effective January 14, 2005, except correspondence for Maintenance Fee payments, Deposit Account Replenishments (see 1.25(c)(4)), and Licensing and Review (see 37 CFR 5.1(c) and 5.2(c)), please address correspondence to be delivered by other delivery services (Federal Express (Fed Ex), UPS, DHL, Laser, Action, Purolater, etc.) as follows:

U.S. Patent and Trademark Office
Customer Window, Mail Stop _____
Randolph Building
Alexandria , VA 22314

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Qi Han whose telephone numbers is (571) 272-7604. The examiner can normally be reached on Monday through Thursday from 9:00 a.m. to 7:00 p.m. If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil, can be reached on (571) 272-7602.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Inquiries regarding the status of submissions relating to an application or questions on the Private PAIR system should be directed to the Electronic Business Center (EBC) at 866-217-9197 (toll-free) or 703-305-3028 between the hours of 6 a.m. and midnight Monday through Friday EST, or by e-mail at: ebc@uspto.gov. For general information about the PAIR system, see <http://pair-direct.uspto.gov>.

QH/qh
June 6, 2005



DAVID D. KNEPPER
PRIMARY EXAMINER